Big Data in Malaysia:
Emerging Sector Profile

BIG DATA Malaysia

2014

Big Data in Malaysia: Emerging Sector Profile 2014

Contact us at sandra@bigdatamalaysia.org, tirath@bigdatamalaysia.org
http://survey.bigdatamalaysia.org

Attribution
Excerpted material from this report can be cited in media coverage and institutional publications. Text excerpts should be attributed to Sandra Hanchard and Tirath Ramdas. Graphs should be attributed in a source line to: Big Data Malaysia 2014

Foreword by Multimedia Development Corporation (MDeC)

Big Data Analytics is a crucial sector in growing Malaysia's digital economy. Multimedia Development Corporation (MDeC) is committed to building a vibrant and dynamic Big Data Analytics industry as a key component of ICT services in Malaysia. The Malaysian ICT Services sub-sector has huge potential growth, with a projected share of 48%[1] in the nation's digital economy in 2020. These ICT services sub-sectors include big data analytics, cloud, mobility, social media and telecommunications. ICT services encompass software development and IT services.

MDeC's strategic intent is for Malaysia to be a hub for Big Data Analytics services, both as a producer and consumer by 2020. Our goal is to catalyse the adoption and usage of Big Data Analytics across all private and public sectors. This reflects the Digital Malaysia initiative

1 *IDC and MDeC Analysis*

which MDeC is driving, to advance the country towards a developed digital economy by 2020. Digital Malaysia aims to create an ecosystem that promotes the pervasive use of digital technology in all aspects of the economy to connect communities globally and interact in real time. This will ultimately result in increased Gross National Income (GNI), enhanced productivity and improved standards of living.

To game change Malaysia's digital landscape, last year MDeC announced the Digital Malaysia 354 (DM354) Roadmap, which encompasses a plan that addresses three ICT areas: access, adoption and usage across five key subsectors: ICT Services, eCommerce, ICT Manufacturing, ICT Trade and ICT Content and Media, targeting four important Digital Malaysia communities. These communities are the B40 group (the lower percent of the Malaysian population in terms of household income), digital entrepreneurs, small to midsized enterprises (SMEs) and youth, whereby elements of Digital Malaysia's initiatives will be proactively embedded in selected programmes and initiatives.

The DM354 initiative will roll out with the ICT services sector which will see the growth of BDA adoption. With the participation of the government and private sectors, BDA adoption will not only reduce costs and raise productivity and competitiveness but also provide a comprehensive and holistic approach that is needed to grow the Big Data ecosystem in Malaysia. The DM354 roadmap was approved by the Prime Minister, Najib Razak last November at the 25th MSC Malaysia Implementation Council Meeting. This council meeting approved the development of Malaysia's Big Data framework, relevant government pilot projects and funding for the private sector.

MDeC and Big Data Malaysia organisations share common goals such as improving Big Data literacies in Malaysia and in helping to connect supply with demand through networking opportunities. MDeC has worked with Big Data Malaysia during the Big Data Week festivals in Kuala Lumpur for the anchor events, "Big Data: Strategy, Practice, and Research" in 2013 and the forthcoming "Big Data: Endless Possibilities" in 2014. Last year we welcomed diverse participants from engineers to executives, across all organisational levels and are looking forward to an even greater response this year. We are delighted to support Big Data Malaysia's survey 2014 report, "Big Data in Malaysia: Emerging Sector Profile". We commend the goal of the report in raising the profile of Big Data initiatives in Malaysia and identifying areas of need and opportunities for growth.

Achieving Digital Malaysia's 2020 aspirational goals is a challenge for the whole nation. Issues of data management and governance, talent development, funding, open data, policies and regulations, and change management will continue to shape MDeC's agenda in supporting the Big Data Analytics industry in Malaysia. We welcome the sharing of knowledge, successes and learnings in supporting each other through this journey.

Chief Executive Officer

Datuk Badlisham Ghazali

Table of Contents

Acknowledgements

The first portion of this report is based on a survey conducted in October 2013. To encourage survey questionnaire completions, lucky draw prizes were offered. We would like to thank the sponsors of these lucky draw prizes:

- Olygen (organisers of the "Big Data World Show" conference series)
- Merlien Institute (organisers of the "Market Research in the Mobile World" conference series)
- Revolution Analytics

We would also like to acknowledge our questionnaire testers for their invaluable feedback prior to the launch of our survey. Our thanks go to Diana Jayne Gonzales (Big Data Philippines), John Berns (Big Data Singapore), Ande Gregson (Big Data Week), Mark Smalley (BrainControl), Daniel Walters (Experian Marketing Services), Jasper Lim (Merlien Institute), Pak Mei Yuet (Multimedia Development Corporation), Luke Lai (Olygen), Sivaramakrishnan Narayanan (Qubole), Andrew Loo Hong Chuang (Taylor's University), Joethi Sahadevan (Taylor's University), Szilard Barany (Teradata), Gopi Kurup (Telekom Malaysia Research & Development), Syahrizal Salleh (Telekom Malaysia Research & Development), Margie Ong (Thoughts In Gear), Rei Lim (Transtel Technology) and Annie Hibberd (VLT).

For assistance with dissemination of our survey questionnaire to their networks and other forms of support, we would like to thank the Multimedia Development Corporation (MDeC), the Association of Banks in Malaysia (ABM), Big Data Singapore, Big Data Philippines, and EmTech Singapore. We would like to thank all respondents for contributing their precious time so that we could have the data needed to conduct the analyses. Furthermore, we would like to thank respondents who shared with us quotable insights, including Jin Chuan Tai (ChrysaSys Consulting) and Greg Whalen (Pivotal) and others who preferred to remain anonymous. The second portion of this report features profiles of Big Data thought leaders in Malaysia. We are very grateful to have been able to interview Heislyc Loh (CoRate), Tom Hogg (Effective Measure), Siim Saarlo (STATSIT), Robin Woo (Western Digital Corporation) and Ian Phoon (from a Malaysian bank) for their time. A note of thanks also to Raju Chellam (Dell Enterprise Solutions) and Elena Tan (JobStreet.com) for participating on a panel dedicated to this report during Big Data Week 2014.

For assistance with editing this report and providing feedback on our survey questionnaire, we owe a huge debt of gratitude to Heoh Chin-Fah (Storage Networking Industry Association Malaysia). Finally, we thank everyone who has contributed to the development of Big Data Malaysia. We value all contributions; simply attending a meetup or posting something interesting to our Facebook wall generates value for the group.

Sincerely,
Sandra & Tirath

Introduction

Big Data has become part of everyday organisational parlance. Increasingly, this awareness is being transformed into practice. Support for harnessing data to amplify capabilities and achieve organisational objectives at practitioner and senior management levels is becoming aligned. A report by Forrester[2] on Big Data adoption in Asia Pacific in 2013-2014 observed a trend across all industries of *'using more types of data, from more sources, to enable timelier better-informed insights'*. Similarly in Malaysia, there is growing cultural acceptance that Big Data can and should enhance decision-making processes, although pathways for adoption are not uniformly understood. Nevertheless, we are now observing a transition phase from curiosity and enthusiasm to buy-in and action across startup, corporate and government organisations. This groundswell of interest fuels the basis of Big Data Malaysia, a networking group for professionals with interest in all things Big Data, including NoSQL, Hadoop, data science, visualisation, business use cases, data governance, open data, and more. Our community has welcomed participation from stakeholders ranging from computer scientists to data journalists, reflecting a broad societal interest in Big Data. Our mission is to encourage high caliber knowledge sharing and to provide a space for professionals with different interests to collaborate. This report grew out of a need to understand in more detail the various networks that Big Data Malaysia helps to connect. In order to support the Big Data ecosystem that we see emerging, we identified critical questions that required further investigation, in particular:

- *What are the opportunities and barriers to Big Data activity in Malaysia?*
- *Who is merely 'interested', versus who is actually committed?*
- *What is the current and future capacity for Big Data talent?*
- *Where are the critical gaps in training and skills?*
- *What are the soft inhibitors, including data access, regulation and perception?*

There are two parts to this report. The first includes results from a questionnaire, while the latter features interviews conducted in-person or via email. In collaboration with various partners, we devised and distributed a questionnaire online across our networks and collected responses during October 2013. In our final sample, we collected responses from 108 individuals over 90 organisations. As our report will show, these viewpoints represent a diversity of organisational stakeholders and industries in the Big Data space. We followed up with interviews of high-profile respondents for richer insight.

---

2 Page 2, Big Data Adoption Trends in Asia Pacific: 2013 to 2014, John Brand and Michael Barnes, Forrester, 2013.

Who should read this report

The objective of this report is to raise the profile of Big Data in Malaysia. We conducted this research because we are passionate about the potential for Big Data to transform the Malaysian economy, provided the right conditions prevail to allow the ecosystem to thrive. If you share this broad view, then this report may be for you.

Although anyone with an interest in Big Data will find something of value in this report, it was designed especially with the following perspectives in mind:

- You are researching how to facilitate engagement with Big Data initiatives across various organisation levels.
- You are preparing to jumpstart your organisations' Big Data initiatives, and are trying to anticipate internal and external challenges.
- You are considering what use-cases you may target for your own Big Data Proof of Concept (POC) projects and are trying to identify what resources you will need.

This report shines a spotlight on the perspectives of senior management and practitioners (e.g. analysts and engineers), across producers and consumers in the Big Data ecosystem. This is the first time we have run the Emerging Sector Profile survey focused on the Malaysian market. Future goals include longitudinal and cross-market analysis to provide market measures and benchmarks. Our scope is limited to presenting a 'grass roots' profile of Big Data in Malaysia based on a thriving, networked community; serving as a foundation to inform readers of this report who may then go on to formulate recommendations that are relevant to their organisation.

Key Findings

About the respondents

- **Interest in Big Data is emerging from both a consumption and production perspective.** Slightly more than 50 percent of respondents claimed to be in an industry vertical other than Information and Communications Technology (ICT), which indicates interest across the broader Malaysian economy beyond any specific sub-sector.

Enabling organisations

- **There is complacency in how managers view the ability of their organisations to leverage data.** Three out of five managers said either they 'Agree' or 'Strongly agree' that their organisation is effectively deriving tangible benefits from their organisational data assets.

- **Anticipated spend for Big Data in 2014 was bullish.** Nearly a third of managers said that they would be increasing their spend on Big Data in 2014 compared to 2013 by more than 25 percent.

- **High outsourcing intent amongst managers in ICT industries suggests technical fragmentation in Big Data industries.** There are opportunities for boutique firms in Malaysia to meet specialist technical needs for global markets.

End uses

- **Customer-centric uses of Big Data dominates prioritisation.** This suggests that interest in Big Data is primarily driven by the pursuit of top-line revenue and business growth, rather than bottom-line factors such as efficiency.

- **Managers need to align the perception of the value of Big Data for both internal and external goals towards a data-centric organisation.** This means recognising both customer-centric and operational end-uses of Big Data.

Skills

- **Differences in desired skills acquisition between ICT and non-ICT firms are likely to produce a sustainable Big Data ecosystem.** However some organisations may participate in the Big Data ecosystem as both producers and consumers, especially organisations in the ICT and marketing services sectors.

- **There is a need to promote the value of fundamental applied math and statistics skills.** These may be overlooked due to the hype surrounding techniques perceived as more 'sophisticated'.

- **Specialised data analysis, modeling, and simulation are the most commonly desired skill.** Intermediate and advanced respondents rated the following five skills as the most desirable:

    - Big and Distributed Data (eg. Hadoop, MapReduce)
    - Algorithms (eg. computational complexity, CS theory)
    - Machine Learning (eg. decision trees, neural nets, SVM, clustering)
    - Back-End Programming (eg. JAVA/Rails/Objective C)
    - Visualisation (eg. statistical graphics, mapping, web-based dataviz)

Capabilities

- **Uncovering patterns in real-time is a strongly desired capability.** 'Real time insights from real-time data streams' and 'Uncovering patterns (e.g. segments, correlations) from multi-structured datasets' consistently placed in the top three in terms of priority.

- **Data visualisation is the number one priority for managers.** Data visualisation is a highly effective way to reveal patterns and trends that computer systems may not yet be able to detect in a fully-automated fashion.

Data sources

- **Most organisations have a single source of data rather than combining internal and external sources.** There are a segment of users that would benefit from external consultation in getting started with the basic goal of identifying data sources.

- **Open data from government sources is needed to support the Big Data ecosystem.** Two out of five respondents said that they either 'Agree' or 'Strongly agree' that having access to some government data does or will create valuable Big Data opportunities for them.

- **There is a general weariness by respondents of bureaucracy and red tape in achieving their Big Data projects.** About one in five respondents said they believe their local legal and regulatory environment to be a hindrance to their Big Data initiatives.

- **Public cloud vendors need to identify and address concerns inhibiting adoption.** This includes privacy, resource cost, and perceptions specific to Malaysian professionals. Only one in four respondents was willing to upload internal data to a public cloud service.

Profiles

- **There are pockets of maturity in the Malaysian Big Data ecosystem.** This includes the use of sophisticated tools and techniques towards niche, high value use cases. Areas of expanding interest we profiled included market research (e.g. brand monitoring, demographic profiling, and detection of purchasing intent), manufacturing (e.g. root cause analysis), finance (e.g. risk management), and many others (including ideas in web applications).

- **Malaysians share the global trend of increasing concern to do with data privacy.** The Personal Data Protection Act 2010 (PDPA) has had an awareness-raising effect.

Seven broad recommendations

- Foster a data-centric organisational culture, where organisational processes, decisions, and objectives are supported by leveraging a variety of data resources.

- Equip your organisation to manage and make decisions based on real-time data, in order to remain competitive.

- Cultivate multi-disciplinary data science teams, consisting of specialists with advanced skills in mathematics and statistics, domain experts, and others.

- Combine data from internal sources within your organisation together with external sources (including public and proprietary third-party sources) to spur innovation.

- Create a 'data dictionary' of data assets relevant to your organisation, to minimise friction for data science projects across your organisation.

- Large organisations, and especially government, should make concerted efforts to release data to the public in machine-consumable form to unlock value hidden within their data assets.

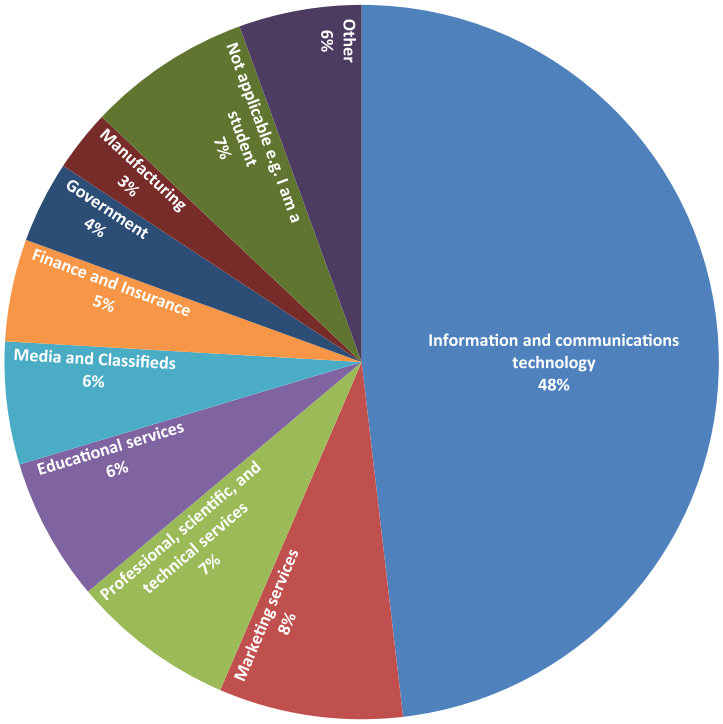
- The PDPA should be frequently reviewed, and revised if necessary, to protect individuals while simultaneously encouraging data projects.

Survey

About the respondents

Substantial interest in Big Data from a consumption perspective as well as a production perspective was indicated by our respondent pool. Slightly more than 50 percent of respondents claimed to be in an industry vertical other than Information and Communications Technology (ICT). While the ICT sector dominated our respondent pool, interest was indicated by other major sectors such as marketing services, professional/scientific/technical services, educational services, and media and classifieds.

Figure 1: What industry vertical is your organisation primarily in?



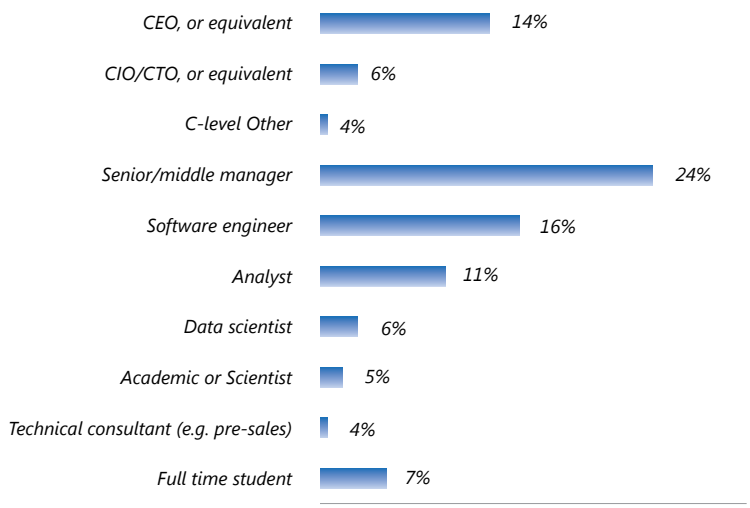
n=108
Source: Big Data malaysia

The survey captured a broad range of professional perspectives at all levels of the organisation. We deliberately did not want to limit our view to executives or purchasing decision makers because our goal was to highlight the state of the overall Big Data ecosystem. Consequently, the respondent pool reflects a balance between decision makers and a class of respondents that we call 'practitioners', who have a 'hands on' role with Big Data.

Overall, 24 percent of respondents were C-level executives (a combination of CEO or equivalent, CIO/CTO or equivalent, and other C-level), a further 24 percent of respondents were

senior/middle management, and 40 percent of respondents were practitioners (comprising software engineers, analysts, data scientists, technical consultants, as well as scientists and academics). It is encouraging that we also received some responses from full-time students, indicating that awareness of Big Data has reached into local universities and colleges.

Figure 2: Job title that is the closest fit to your current role:



n=108
Other = 5
Source: Big Data Malaysia

The respondent profile in terms of organisation size was also reasonably balanced. For respondents who fell into the 'practitioner' or 'senior/middle manager' role categories, the organisations they represented ranged in size between very small organisations (including organisations with less than 5 employees as well as the self-employed) and large organisations (with over 1000 employees). However, our C-level respondent pool was heavily skewed towards small and medium sized organisations, with almost 90 percent of C-level respondents being with organisations of less than 75 employees.

Create a 'data dictionary' of data assets relevant to your organisation, to minimise friction for data science projects across your organisation.

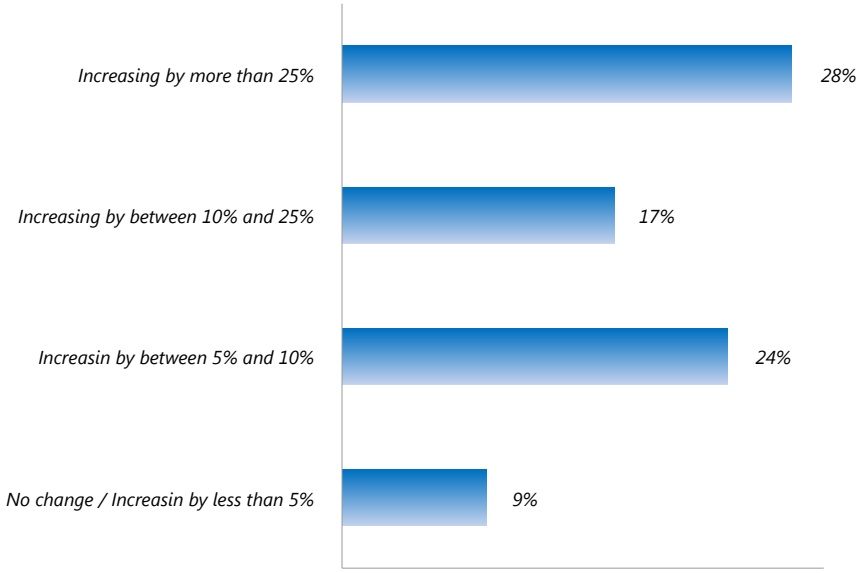
Enabling organisations

Big Data approaches are becoming infused into organisational cultures in Malaysia, reflected by resources being dedicated to Big Data projects. We asked respondents with managerial responsibilities to provide a 'temperature' of their Big Data readiness and resourcing intent for Big Data in 2014. Overall, we found that aspirations were high amongst managers, but actual allocation to personnel was conservative. The results were nevertheless encouraging in managers matching enthusiasm for Big Data with resource commitment.

There was some complacency in how respondents viewed the ability of their organisations to leverage data. Three out of five managers said either they 'Agree' or 'Strongly agree' that their organisation is effectively deriving tangible benefits from their organisational data assets. Just under a third of managers were neutral, while just over one in ten managers said they either 'Disagree' or 'Strongly disagree'. However, organisations need to ensure they are not over-estimating their capacity to leverage their data. A discussion with Greg Whalen, Pivotal's Chief Data Scientist, Asia Pacific and Japan, revealed that many organisations do not have a 'data dictionary.' Data scientists use this 'inventory' of enterprise data to identify new correlations that provide return on investment through creative use of existing information. While some managers might believe they are already using their data assets effectively, newer approaches in data mining, exploratory visualisation and comparison across external data sets may shed new light on unknown opportunities.

Anticipated spend for Big Data in 2014 was bullish. Nearly a third of managers said that they would be increasing their spend on Big Data in 2014 compared to 2013 by more than 25 percent. This large increase in investment suggests a small base of comparison, supporting arguments that we are experiencing a turning point in Big Data in Malaysia from 'awareness to implementation'. Only 9 percent of managers said they were increasing their Big Data budget by less than 5 percent. The remaining bulk of managers (41 percent) said they were increasing their budget from 5 to 25 percent. There was a similar spread in spending intentions between managers in ICT and non-ICT industries. In terms of increasing headcount specifically around Big Data initiatives, most respondents said they expected to grow their staff in the 4-10 range, with few anticipating growth of more than 20 staff in the coming year.

Figure 3: How do you expect your spend on Big Data to change in 2014 compared to 2013?



n=75
Don't know / prefer not to say = 16
Source: Big Data Malaysia

High outsourcing intent amongst managers in ICT industries suggested technical fragmentation in Big Data industries. While 48 percent of non-ICT managers said they were either moderately or extremely willing to outsource high-skills tasks in their Big Data initiatives to external consultants, there was greater willingness in ICT industries, with a comparative number of managers accounting for 69 percent. This suggests increasing technical fragmentation in Big Data. There are opportunities for boutique firms in Malaysia to meet specialist technical needs. Local firms can nurture local talent and export to global ICT markets

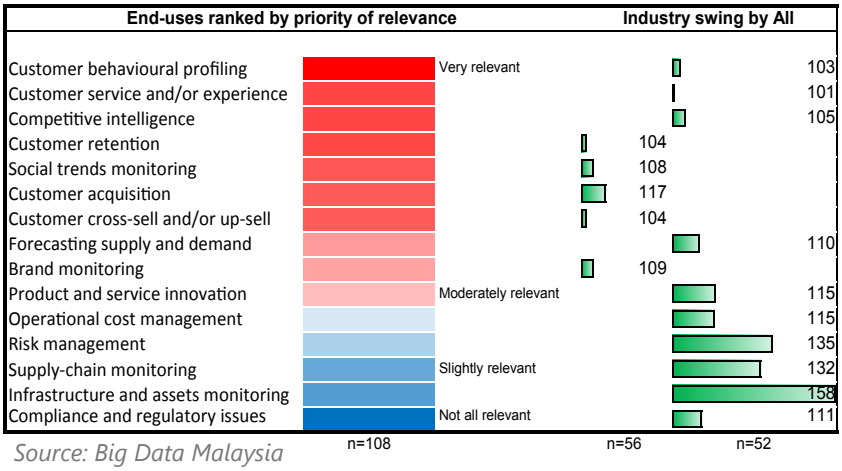
Foster a data-centric organisational culture, where organisational processes, decisions, and objectives are supported by leveraging a variety of data resources.

End uses

Applications of Big Data may be relevant to every organisational process and objective, whether internally or externally focused. While some uses of Big Data are clearly understood and valued, others are still emerging. Greg Whalen, of Big Data platform vendor, Pivotal, shared with us successful use cases ranging from 'identifying drivers of manufacturing quality to detecting malware outbreaks to predicting customer churn to identifying at-risk patients.' While applications might be relevant to specific domains, sharing case studies across industries can highlight creative processes in extending data analysis. Ultimately, there is a need by organisations to determine how Big Data strategies align with their organisational values and objectives.

Customer-centric uses of Big Data dominated prioritisation by respondents. Customer behavioural profiling was the most highly rated end-use in terms of relevance. Improving customer service and experience, retention, acquisition, cross-sell and upsell were also identified as important uses of Big Data. Respondents also valued Big Data for 'competitive intelligence' in benchmarking against their industry. The high ranking of social trends monitoring highlighted the general appeal of social data. Of less relevance to respondents were uses rated to compliance and regulatory issues, infrastructure and assets monitoring, supply-chain monitoring, risk management and operational cost management. This suggests that currently, interest in Big Data is primarily driven by the pursuit of top-line revenue and business growth, rather than bottom-line factors such as efficiency.

Figure 4: How relevant are the following end-uses for Big Data capabilities from your own perspective?



Source: Big Data Malaysia

There were differences in prioritisation between respondents in ICT and non-ICT industries. Non-ICT respondents were more likely to value social trends and brand monitoring, Conversely, ICT respondents exhibited a swing towards internal operational end-uses. The greatest differences between the two groups of respondents were in infrastructure and assets monitoring, risk management and supply-chain monitoring. ICT respondents also

gave a higher prioritisation to product and service innovation. There is a need for non-technical businesses to appreciate how data-driven feedback can improve products and services for competitive advantage. Differences in prioritisation should prompt stakeholders formulating uses cases of Big Data to consider whether they are too internally or externally focused; there is an opportunity to think outside-the-box for applications and to learn from other industries. Big Data vendors may choose to tailor products to specialised organisational functions.

Managers need to align the perception of the value of Big Data for both internal and external goals towards a data-centric organisation. For example, data-driven approaches in change management, product and service development and customer building may be complemented with data-driven approaches of increasing operational efficiency. In their perception of Big Data uses, managers were more likely than practitioners to value risk management, customer retention, customer acquisition, product and service innovation, customer behavioural profiling and forecasting supply and demand. Practitioners had a stronger internal operational focus, being more likely to prioritise supply-chain monitoring and infrastructure and assets monitoring. The mantra of reducing cost while increasing value should be at the core of applying Big Data in organisations.

Cultivate multi-disciplinary data science teams, consisting of specialists with advanced skills in mathematics and statistics, domain experts, and others.

Skills

One of the most frequently cited obstacles to delivering Big Data initiatives is the difficulty in recruiting the required skills. We presented respondents with a list of skills[3] and asked them to score each skill according to need from their perspective. Each response scored on a 5 point scale, ranging from "No need" to "Critical need", then scores aggregated to produce the following heatmap:

Figure 5: To deliver your Big Data initiatives, how much need is there to recruit the following skills?

	ICT only	All respondents	Non-ICT only
Specialised data analysis, modeling, and simulation (e.g. operations research, machine learning, etc.)	3	Priority 1	Priority 1
Distributed systems (e.g. Hadoop) deployment and/or administration	Priority 1	2	2
Fundamental computer science and/or software engineering	4	3	3
Industry-specific/domain knowledge	2	4	4
Applied math and/or statistics	5	5	6
Web/mobile development and/or visualization	7	6	5
Research experience from any quantitative discipline	6	7	7
Business (strategy, marketing, product development, etc.)	8	8	8
Hardware/sensor design	9	9	9
	n=52	n=108	n=56

Source: Big Data Malaysia

The most commonly desired skill across the combined respondent pool was specialised data analysis, modeling, and simulation. This was followed by distributed systems, deployment and / or administration, and fundamental computer science and software engineering. There was consensus across segments that pure business skills (distinct from domain knowledge) was of low importance, but the least relevant skillset was hardware/sensor design. We included this skillset in our list as a proxy for the "Internet-of-Things" concept that is often seen as closely related to Big Data.

3 Our list of skills was derived in part from the Skills List presented in "Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work" by Harris, Murphy, and Vaisman, 2013. We grouped together several related items from their list into composite items in our list.

There is nuance in skills priorities between ICT and other industries. The overall respondent pool cited specialised data analysis, modeling, and simulation as their most pressing need, with distributed systems (e.g. Hadoop) coming in a close second. However, when segmenting by industry (ICT vs all other industries), we see skews in preferences:

- Specialised data analysis, modeling, and simulation:
  - 1st priority by non-ICT respondents.
  - 3rd priority by ICT respondents.

- Distributed systems deployment and/or administration:
  - Distant 2nd priority by non-ICT respondents.
  - 1st priority by ICT respondents.

- Industry-specific/domain knowledge:
  - 4th priority by non-ICT respondents.
  - 2nd priority by ICT respondents.

The difference in domain knowledge may be expected (in particular if ICT respondents aim to provide services to non-ICT consumers, in which case they would need to acquire target domain knowledge to provide adequate service), but the difference in distributed systems versus specialised data analysis may point to different aspirations by the respondent groups.

Differences in skills acquisition between ICT and non-ICT firms is likely to produce co-dependencies that shape a sustainable Big Data ecosystem. Broadly speaking, if we may assume that a significant portion of ICT respondents are participating in the Big Data ecosystem as solutions providers to non-ICT respondents and that only a small portion of non-ICT respondents are participating in the Big Data ecosystems as providers, then our findings may indicate the emergence of a model of engagement with ICT players providing Big Data infrastructure capabilities and solutions, with some degree of industry customisation. Meanwhile, non-ICT consumers are developing deep analytical capabilities pertinent to their business needs on top of the provided infrastructure, notwithstanding the role of small boutique ICT and analytics consultancies that can fill gaps wherever opportunities arise. Some organisations may participate in the Big Data ecosystem as both producers and consumers, especially organisations in the ICT and marketing services sectors.

There is a need to promote the value of applied math and statistics skills for Big Data analytics. It is noteworthy – and likely a point for caution – that respondents rated specialised data analysis skills so highly but were lukewarm to applied math or statistics skills. It is possible that the respondent pool have already developed significant practical expertise in applied math and statistics, but another likely scenario is that there exists a lack of understanding of the utility and value of applied math

and statistics to Big Data analytics. The hype around more novel approaches such as machine learning may be attracting disproportionate mindshare over applied math and statistics, which respondents may have perceived as "basic"[4] .

The breadth of our respondent pool gives rise to a big range in technical knowledge, which might lead to concerns that some respondents may have not had sufficient understanding to properly assess their need for some specific technical skills. To account for this possibility, we followed up with additional data collection. Firstly, we asked respondents, *"How would you assess your own degree of technical understanding of Big Data?"*[5] . Respondents chose one of the following options: 'Total beginner', 'Technical details are not relevant to my job role", 'Intermediate', and 'Advanced.' Respondents who answered 'Intermediate' or 'Advanced' (56 percent of our total respondent pool) were then presented with an expanded skills list[6] and asked to select only their first, second, and third skills priorities. The responses for each skill were weighted by priority and the resulting scores aggregated. These results from skilled respondents reinforce our earlier findings:

- ICT respondents are skewed towards distributed systems skills.
- Non-ICT respondents are skewed towards advanced algorithms.
- There is lukewarm interest in math and statistical methods.

An additional finding is that despite the overall lukewarm interest in statistical methods, there is one bright spot: temporal statistics. This points to interest in time series analysis, which is an important method applicable to many areas.

Despite differences between ICT and non-ICT perspectives amongst intermediate and advanced respondents, it is clear that there is significant shared interest in the top five skills:

- Big and Distributed Data (eg. Hadoop, MapReduce)
- Algorithms (eg. computational complexity, CS theory)
- Machine Learning (eg. decision trees, neural nets, SVM, clustering)
- Back-End Programming (eg. JAVA/Rails/Objective C)
- Visualisation (eg. statistical graphics, mapping, web-based dataviz)

Targeted training in these areas, especially if offered in combination with each other, is likely to see widespread interest and provide

4 In the "Profiles" section of this report, we will show that several thought leaders suggest that skills such as statistics and a scientific background are of significant value to Big Data initiatives, contradicting respondent perceptions.

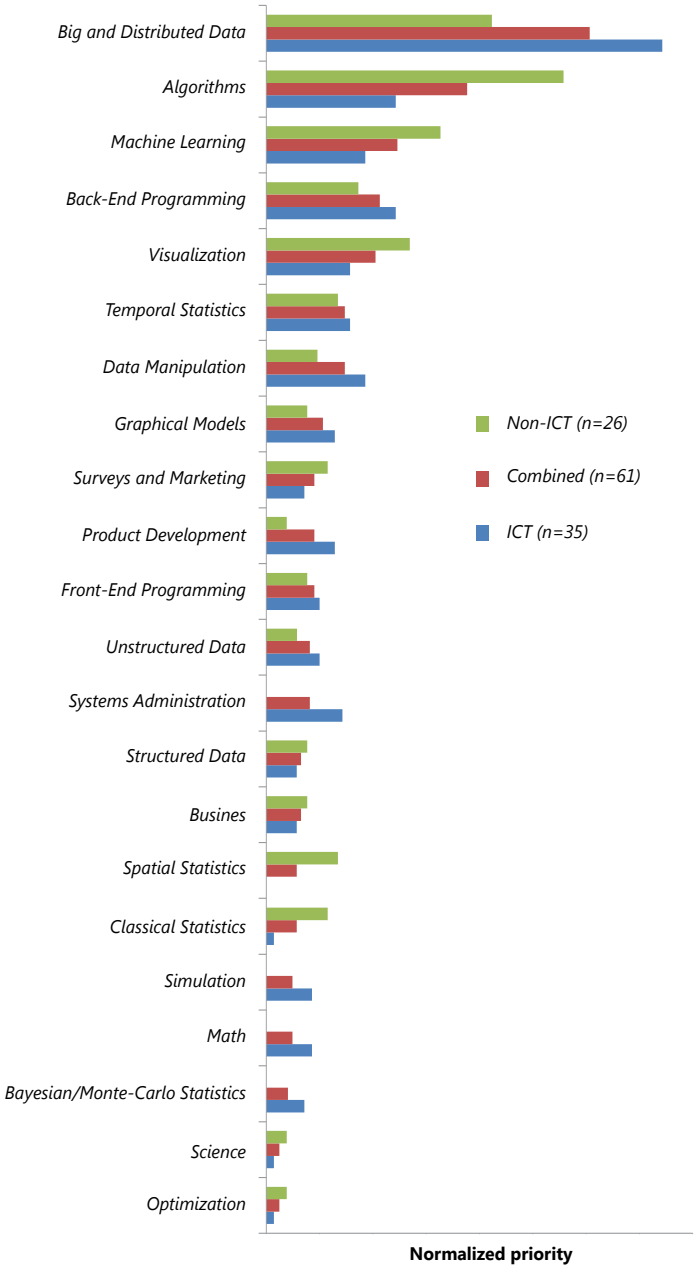
5 This self-rating question only served as a basis to drill down into follow-up questioning; in no way do we suggest it may be used as an accurate gauge of the extent of Big Data technical knowledge in Malaysia.

6 Here we used the full Skills List from "Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work" by Harris, Murphy, and Vaisman, 2013.

a major boost to the development of the Big Data ecosystem in Malaysia.

Figure 6: If you (or your employee/service provider) could only develop one skill, what should it be?

Respondents presented with a skills list and asked to select their first priority, second priority, and third priority only. Responses for each skill scored according to priority, then aggregated.



Source: Big Data Malaysia

Equip your organisation to manage and make decisions based on real-time data, in order to remain competitive.

Capabilities

This section highlights specific capabilities that respondents seek in their Big Data solutions. Capabilities are desired characteristics and features that may cut across industry segments, be applicable to multiple end-uses and/or rely on some combination of skills to produce and consume. As such, we investigate capabilities as a separate, though not necessarily independent issue. Some capabilities are specific to certain types of data, but we still regard these as unique capabilities rather than data source issues because simply having access to the data does not produce desired outcomes – it is the handling of the data that creates value.

A meaningful opinion on this issue requires some degree of technical knowledge, so we only included respondents who answered either 'Intermediate' or 'Advanced' to the question, "*How would you assess your own degree of technical understanding of Big Data?*[7]" (56 percent of our respondent pool). In addition to drilling into ICT vs other industry segments, here we also investigate the set of respondents who are "Managers", i.e. respondents who answered 'Yes' to the question, '*Does your role in your organisation include managerial/leadership responsibilities?*' (69 percent of our respondent pool). The subset of our respondent pool that comprise managers with at least 'Intermediate' technical understanding of Big Data make up 43 percent of our total respondent pool.

Uncovering patterns in real-time is a strongly desired capability. Despite differences in perspective in the different segments studied, 'Real time insights from real-time data streams' and 'Uncovering patterns (e.g. segments, correlations) from multi-structured datasets' consistently placed in the top 3 in terms of priority. These priorities were especially strong for ICT respondents, who went on to indicate 'Data discovery and exploration across many data sources' as their third priority. Taken in combination, these are capabilities that are especially related to Big Data infrastructure, once again illustrating an apparent inclination for ICT respondents to focus on supplying capabilities for external consumption, rather than for their own internal consumption.

7 Respondents choose one of the following options: 'Total beginner", 'Technical details are not relevant to my job role", 'Intermediate", and 'Advanced".

Figure 7: Which of the following Big Data capabilities are relevant to you?

	ICT only	All respondents	Non-ICT only	Managers only
Real-time insights from real-time data streams	2	Priority 1	Priority 1	3
Uncovering patterns (e.g. segments, correlations) from multi-structured datasets	Priority 1	2	3	2
Visualizing/presenting insights	4	3	2	Priority 1
Data discovery and exploration across many data sources	3	4	4	4
Statistical analysis on big working data sets (>100GB)	5	5	5	5
Automated decision making	7	6	7	6
Machine-generated data (e.g. log files, periodic diagnostics)	6	7	9	8
Content and sentiment from online media (e.g. social media)	9	8	6	9
Efficiently and safely storing large data sets on infrastructure controlled by my organization	8	9	10	7
Image, video, and audio data	10	10	8	10
Physical sensor networks (e.g. "Internet of Things")	11	11	11	11
	n=35	n=61	n=26	n=46

Note: Each response scored on a 5 point scale, ranging from "Not at all relevant" to "Very relevant", then scores aggregated to produce the above heatmap.

Source: Big Data Malaysia

Data visualisation is the number one priority for managers. This capability was also strongly desired by non-ICT respondents. This points to a consumption need; non-ICT and especially manager respondents may prioritise Big Data capabilities to support their decision making processes, and visualisation of data is a highly effective way to reveal patterns and trends that computer systems may not yet be able to detect in a fully-automated fashion. In the overall respondent pool, visualisation was only the third priority because it was pulled down by ICT respondents, who only placed it as the fourth priority.

Once again, there is no interest in "Internet of Things" (IoT). Despite the opportunities for linking the two worlds (with IoT serving as a data source and Big Data producing value out of the raw data) it seems these fields, which are still relatively nascent in Malaysia, are yet to instigate much interest in combination. Even within the ICT segment of our respondent pool, there is virtually no interest in IoT, which suggests that even the supply-side is not yet ready to deliver these capabilities to the Malaysian market. As the Big Data and IoT ecosystems continue to evolve in Malaysia we may see this lack of interest give way to interesting startup initiatives. In the meantime, awareness raising initiatives should focus on this link.

Combine data from internal sources within your organisation together with external sources (including public and proprietary third-party sources) to spur innovation.

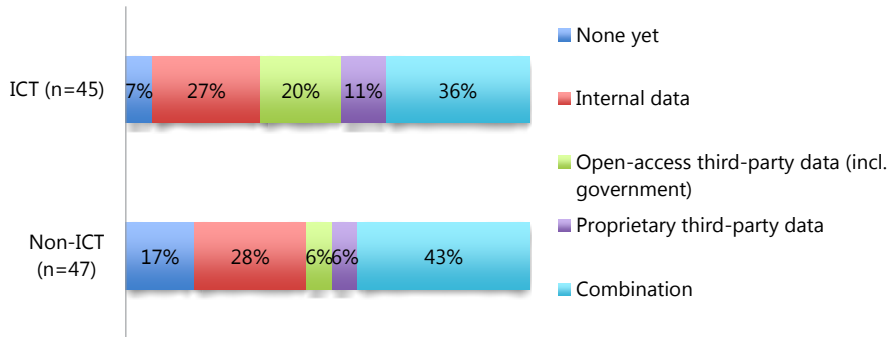
Large organisations should make concerted efforts to release data to the public in machine-consumable form to unlock value hidden within their data assets.

Data sources

Data is the new oil. Dimensions of data quality and value are increasingly as important as volume, velocity and variety. The wider social and business impact of Big Data initiatives is dependent on the quality of data and information on which they are based. Data environments whether open or proprietary are becoming increasingly complex as new business models emerge. Demand for open data for both private and public sector use is increasing for endeavours such as data journalism. In Malaysia, complex social problems such as traffic congestion or local crime could be addressed by open data and crowdsourcing approaches, in both diagnosing problems and in providing feedback mechanisms from citizens. There are many issues related to data sources that we identified during our survey. These include use of internal and external data, access to open government data, regulatory issues and appetite for using third-party infrastructure (e.g. cloud services) for data management. The sourcing and management of data requires consideration of complex relationships and issues, including strategic partnerships, security, and distribution, to name a few.

Most organisations have a single source of data rather than combining internal and external sources. ICT respondents were more likely than non-ICT respondents to use open access (e.g. government) and proprietary third party data as the sole source for Big Data initiatives. Non-ICT respondents were polarised in terms of data resourcing where 43 percent said they were using a combination of sources, while 17 percent of respondents were not using any data for Big Data initiatives. There are a segment of users that would benefit from external consultation in getting started with the basic goal of identifying data sources. Respondents who were likely to highly-value customer behavioural profiling were dependent on third party data such as social media.

Figure 8: My organisation uses sources for Big Data initiatives primarily from the following:



'Don't know/prefer not to say' = 16
Source: Big Data Malaysia

Open data from government sources is needed to support the Big Data ecosystem. Two out of five respondents said that they either 'Agree' or 'Strongly agree' that having access to some government data does or will create valuable Big Data opportunities for them. Just under half of respondents were neutral, while 1 in 10 respondents said they either 'Disagree' or 'Strongly disagree'. The large number of neutral respondents suggests more education is required around how open data can be used for Big Data projects. Respondents identified a number of uses of government data. Topics that were frequently cited included:

- Demographic and socioeconomic data
- Population (online and offline)
- Crime; Border security and migration
- Public and community services (utilities, health, education)
- Location (by utility); Geographic Information Systems (GIS)
- Financial; credit
- Weather

By far the greatest need was demographic and socio-economic data. One respondent, head of decision science at a multinational bank, said that "*In order for us to understand the needs of Malaysians, statistical data from the population census is important to identify correlations with internal behavioural data.*" There is a need for government agencies to share data multilaterally for social problems that have many dimensions. For example, access to health care is also a function of transportation, location and income. Ministries can utilise internal data management for service optimisation. Jin Chuan Tai, Director of ChrysaSys Consulting said that "*Government has many different sets of survey data, collected from various sources. For instance, Ministry of Health data can be sourced from private and general hospitals, clinics or consultancies. Big Data offers a mechanism to speed up consolidation of all this information, without any processing delays to configure each and every source.*"

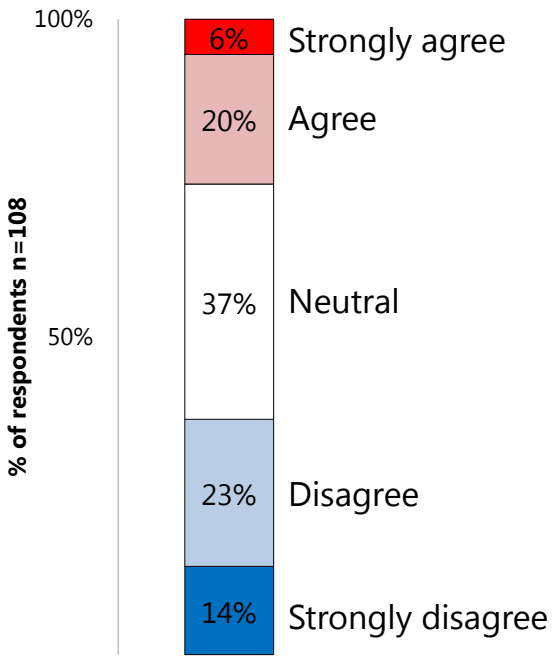
There was a general weariness by respondents of bureaucracy and red tape in achieving their Big Data projects. About one in five respondents said they believe their local legal and regulatory environment to be a hindrance to their Big Data initiatives, compared to 15 percent who did not. Most respondents (47 percent) were neutral. A further 15 percent said 'don't know' clearly identifying need for further education on emerging regulatory trends in the era of Big Data. The most common hindrance cited was the Personal Data Protection Act 2010 (PDPA), gazetted in Malaysia in late 2013[8]. This was a common response highlighting the need for balancing consumer protection with business needs, and better education specifically on the PDPA. Other issues of concern to respondents included data compliance, data risk and loss of data.

8 The hindrance posed by the PDPA on Big Data initiatives may be more in terms of perception, rather than actual legal definition. In the "Profiles" section of this report, we will show that one industry thought leader believes that the PDPA is actually beneficial, by providing reassurance to the public that their data is protected with a legal framework in place.

The PDPA should be frequently reviewed, and revised if necessary, to protect individuals while simultaneously encouraging data projects.

Public cloud vendors need to identify and address concerns – privacy, resource cost, and perceptions specific to Malaysian professionals. Only one in four respondents was willing to upload internal data to a public cloud service. While 37 percent of respondents said they either 'Disagree' or 'Strongly disagree' with the statement, 'I am willing to upload my internal data to third-party infrastructure (e.g. public cloud'), just over a third of respondents were neutral, indicating that there are a large number of fence-sitters that have yet to be swayed either way. Respondents who were more likely to invest in Big Data initiatives were also much more likely to be willing to use public cloud services, suggesting that maturity around data use and Big Data literacy reduces the anxiety around uploading data to third-party infrastructure.

Figure 9: I am willing to upload my internal data to third-party infrastructure (e.g. a public cloud)



Source: Big Data Malaysia

Profiles

Overview

The previous section presented an aggregate view of the state of Big Data activity in Malaysia. In this section we present a closer look, from the perspective of five industry leaders:

- Heislyc Loh, Founder and Chief Executive Officer, CoRate
- Tom Hogg, Commercial Director, Effective Measure
- Siim Saarlo, Chief Technology Officer, STATSIT
- Robin Woo, Senior IT Manager, Western Digital Corporation
- Ian Phoon, Data Scientist, a Malaysian bank

Our objective in selecting these interviewees was to strike a balance in backgrounds in end uses and to highlight diverse industries. We have representatives from finance, information and communication technology, manufacturing, marketing and social media; across roles in Commercial, Digital Marketing and Product departments. The areas of activity and expanding interest in Big Data analytics we profiled included market research (e.g. brand monitoring, demographic profiling, and detection of purchasing intent), manufacturing (e.g. root cause analysis), finance (e.g. risk management), and many others (including ideas in web applications). Some of these use cases are niche interests and specific to certain industries, but may nevertheless be of high value and high impact.

The initiatives described by our interviewees point to the existence of pockets of maturity in the Malaysian Big Data ecosystem. Sophisticated tools and techniques – such as the application of natural language processing, machine learning, full text search and analysis, statistics, social science, data visualisation, and data governance as well as general data quality measures – are being applied towards tasks such as semantic analysis, defect root-cause-analysis, financial early-warning systems, and general predictive analysis solutions. In some situations there is already recognition of the need for specific Big Data return on investment measures. One respondent referred to the conceptual idea of a 'value to bit' ratio, which attempts to capture the objective of increasing the value of data assets, while reducing costs associated with data collection, retention, and management.

The organisations represented here are leveraging a wide variety of data sources, including social media, website traffic and browser interactions, traditional surveys, external third-party media (such as news reports), and others. While in most cases many options for data collection exist, several respondents expressed the view that the Malaysian government should do better in terms of releasing government data, for instance by making available regular census information through promotion to organisations that would be benefit from this data.

Beyond opening up government data, our respondents raised several other points about the prospects of Big Data analytics as a significant part of the Malaysian economy. Firstly, the issue of talent availability was highlighted. Exacerbating this issue, there is a perception that relevant training is relatively lacking in Malaysia (e.g. as compared to Singapore). Until a local pool of skilled specialists can be grown, it may continue to be necessary to outsource the highest-value tasks to specialists elsewhere. There are different opinions on the precise requirements and priorities for data scientists, but a strong working knowledge of statistics seems to be on everyone's wish list. From the Big Data consumption perspective, it is pertinent that the issue of privacy is receiving significant interest in Malaysia. There is a global trend of increasing concern to do with data privacy, and our respondents indicated that Malaysians share these concerns. The Personal Data Protection Act 2010 (PDPA), beyond its legal definition, has had an awareness-raising effect.

As for issues unique to startup ventures, despite a favourable view of Big Data initiatives by funding organisations, the usual factors impacting all IT startups in Malaysia (in particular access to post-seed-level-funding, and exit opportunities) remain. These issues are made more acute by the longer runway required to achieve sufficient scale – to produce sufficient data – such that Big Data analytics can be meaningfully exploited. Some of our respondents here share views based on substantial regional or global experience as a point of comparison to their successes and challenges in the local landscape. All are dedicated to seeing Big Data succeed in Malaysia.

"We're confident our vision will be widely accepted by people who need to deal with large amounts of information."

Heislyc Loh, CoRate

CoRate is an early stage startup that aims to help people consume information on the Internet more effectively, by providing a smart article collection platform - like pocket with automatic categorisation, organisation and summarisation capabilities. These functions will be developed through a combination of crowd sourcing and data analytics. CoRate founder, Heislyc Loh is confident that their vision will be widely accepted by people who need to deal with large amounts of information, ranging from business analysts and researchers to students. He calls this target group of people, 'infoholics.'

End uses and data sources

Going beyond social bookmarking, the platform is an ambitious yet classic example of Big Data, with a roadmap to utilise natural language processing, machine learning, and other elements. It is envisioned that the platform will leverage three major data source categories: data and metadata within Internet articles themselves, data captured on user interaction with the article (e.g. browser extensions will allow users to highlight interesting passages), and interactions and relationships between users (i.e. social data).

Loh recognises that this vision will be challenging to realise in the best of circumstances, but is nevertheless optimistic. Since late 2013 a very early-stage piece of the product has been released in the form of a Chrome web browser extension, but by-and-large development work only commenced at the start of 2014, and building-up of a team in Kuala Lumpur is ongoing. The current team comprises mostly front-end developers with an emphasis on User Experience (UX) as CoRate tries to achieve end-user traction. At this early stage their technology stack consist of web app mainstays MongoDB, Ruby-on-Rails and Javascript.

Malaysian environment and skills recruitment

Loh believes that funding in Malaysia is not an issue at the early stages, in part due to the availability of grants from various government agencies, and Big Data initiatives in particular are viewed favourably. However, in the medium-to-long-term CoRate has its sights on Silicon Valley. This is primarily because despite the availability of early stage funding, there is very little high-capital investment (e.g. Series B and beyond) available in Malaysia. Furthermore, given CoRate's consumer-Internet focus, exit opportunities are relatively thin outside of Silicon Valley.

He goes on to acknowledge that beyond the initial phases, realising CoRate's vision will require high-end specialised skills, and finding individuals in Malaysia with relevant experience would prove challenging. Nevertheless, he believes that in the long term these capabilities can be developed at home because "there are smart people here". To jumpstart the development process he plans to recruit specialists from overseas (primarily from San Fransisco, where he spent a few weeks in late 2013 developing the CoRate concept) with the needed skills.

"We need to showcase the power of the local audience... There's a place for everybody, and room for everyone to operate"

Tom Hogg, Effective Measure

Tom Hogg is the Commercial Director for Effective Measure in Malaysia. Effective Measure is a leading provider of digital audience, brand and advertising effectiveness measurement and targeting solutions, bringing best practice online measurement data to premium publishers, agencies, networks, advertisers and researchers. Effective Measure's client base in primarily made up of publishers, agencies and networks, and they are increasingly working with advertisers and insight agencies. They are headquartered in Melbourne, Australia where they were founded and developed their proprietary technology, which now provides online measurement solutions globally. In South East Asia, Effective Measure operates in Malaysia, Singapore, Thailand, Vietnam, Philippines, Hong Kong and Indonesia. Top advertising clients globally, based on digital spend, are typically in finance, automotive, FMCG (food and beverage, homecare, health and beauty), as well as travel and hospitality. In Malaysia, Effective Measure covers website traffic from 38 million unique browsers each month on its network (including users with multiple devices). They currently offer 70,000 demographic profiles over the online population nearing 20 million people in Malaysia.

Measuring online audiences

In terms of the Big Data ecosystem in Malaysia, Effective Measure provides detailed demographic and psychographic insights into website audiences, reflected by Malaysian internet usage at a census level. One of their core objectives is to help publishers and networks validate who their audiences are, so they can pitch to agencies and advertisers about the kinds of people on their website. A main challenge for local publishers in markets like Malaysia is the competitive influence of global operations, who have large audience reach. Hogg emphasises that it is very important to "*help local publishers have a trusted and transparent source of data that can be shared with agencies and advertisers. We need to showcase the power of the local audience, in addition to the audiences visiting these multinational giants. There's a place for everybody, and room for everyone to operate; the concern for local publishers is that more and more of the AdEx [advertising spend] is leaving local markets, with the risk that local publishers lose their revenues, which ultimately means that local content is non-sustainable.*"

There have been vast changes in media measurement with the advent of Big Data. Hogg notes that online environments offer easier access to much larger data sets, whereas previously advertising decisions were made on small subsets of the population. An important development in online measurement is the ability to compare the impact of advertising on members of a population who have been 'exposed' to a digital campaign versus members who have not. In post-campaign studies, clients can link impacts back to core objectives such as consumer brand advocacy, uplift in brand awareness, purchase intent and a range of insights based on different business outcomes. Because more advertising spend is going online, validation is required by third-party online measurement providers to justify budgets.

Effective Measure presents aggregated, anonymous behavioural data based on groups of people through a dashboard client service. Hogg observes that most people "aren't experts in statistics, so it's really important to make the data easy for people to understand". Data visualisation is increasingly important in how clients consume data. The evolution of infographics in Big Data follows general consumption trends for media i.e. interactive, visually stimulating and appealing. In Malaysia, Hogg has observed an increase in requests by clients for key summary points to be pulled out in addition to their dashboard service.

Regulatory and Malaysian business environment

Hogg says that Effective Measure has observed an increase in sensitivity by Malaysian respondents to some types of survey questions. Hogg believes this is part of a global trend of

increasing consumer concern with data privacy. In addition, the crafting of the Personal Data Protection Act (PDPA) in Malaysia has heightened awareness around data privacy; however, Hogg believes the PDPA is likely to be ultimately beneficial in helping people to understand that their data is protected with a legal framework in place. Government can support the Big Data industry in Malaysia by providing regular census information, insights into online behaviour and information on infrastructure for the purposes of corroboration. The more government data that can be accessed, the better for the entire digital community. In terms of comparing the data analytics market in Malaysia versus the rest of South East Asia, Hogg doesn't see a tremendous difference, with the exception of Singapore. Hogg observes an equal interest in data analytics in Malaysia as in other markets; however, there is a range of data literacy within Malaysia. Education is required on understanding statistical methodologies in how panels are used to estimate population sizes; this can help both publishers and advertisers understand Big Data methods better.

"There is hardly a profession in IT that would not benefit from some data awareness."

Siim Saarlo, STATSIT

Siim Saarlo is the Chief Technology Officer of STATSIT, a social media marketing technology company. STATSIT works with brands and agencies in the Asia Pacific region to help them increase advocacy and sales, to identify influences and to prove the value of social media activities. In general, STATISIT extracts insights from social media conversations. They define relevant subsets of conversations and conversationalists for customers and projects. Over the resulting stream of social data, they perform analysis and produce insights in various forms. This includes basic volume monitoring, tailored reports and action steps, influencer profiling and predictions on content demand.

Practices

STATSIT relies on data science practices and the statistical capabilities of social scientists, as well as machine learning algorithms. New approaches and algorithms get tested first on internal and one-off projects before reaching end products. Due to their normal subject matter, STATSIT utilise natural language processing (NLP) to some extent. However, STATSIT avoid more complex NLP algorithms because these are usually language

specific while their main products cover a variety of languages. Their methodologies for extracting insights from social data depend on project goals. They rely on both standardised quantitative elements and some qualitative analysis leading to standalone products e.g. influencers discovery and analysis. Qualitative analysis mainly relies on analyst work and is often iterative, but is supported with different analytical tools, such as full text search, entity and pattern detection etc.

Infrastructure

Saarlo believes that organisations have an opportunity to optimise their resource usage in terms of infrastructure. He comments that riding the big data wave often makes organisations collect more than they analyse, leaving them with a lower 'value to bit ratio'. Recognising this challenge, STATISIT are looking for options to change their data collection, processing and usage processes to cut down costs for infrastructure and at the same time increase the value of generated insights. This is not so much with the aim for cutting cost, but to focus more on value creation before data collection. This might mean reviewing their agreements with infrastructure providers, employing new and more effective technologies, being smart in reducing the amount of 'active data' and focusing analytical efforts etc. In summary, their objective is to be more effective and smart with both technology and data usage.

Data environments

STATSIT works mostly with data that is public and available for free. From the perspective of a product developer, an ongoing challenge is when platforms restrict access to data as business models mature. Saarlo believes that general data aggregator companies will get more dominant over time, and that data collection and data analysis will become increasingly siloed activities. Saarlo predicts greater specialisation in big data with more companies tackling very specific domain problems in niche fields. Additionally, he sees an increasing utilisation of open source components versus approaches that port general machine learning algorithms into completely proprietary solution stacks.

In order to support data intensive endeavours, Saarlo believes that governments and organisations should make data available as much as reasonable, keeping in mind an audience of curious and innovative developers. He says that as data science is fairly well popularised in professional circles, students should also get taste of the hype, which would naturally increase interest in subjects like maths and statistics. Saarlo says that there is "hardly a profession in IT field that would not benefit from some data awareness".

"One must know how to ask the right questions."

Robin Woo, Western Digital Corporation

Western Digital Corporation is one of the largest computer storage manufacturers in the world, and is one of the most important manufacturers in Malaysia. Robin Woo, Senior IT Manager at Western Digital Malaysia, is responsible for leading datacentre operations management and systems administrations at all manufacturing sites worldwide. He is also responsible for leading web development for general IT applications globally, and for leading the company's IT roadmap strategy globally.

Scope and context

In Woo's view, Big Data and Analytics are inseparable because ultimately it is what you do with the data that counts. In his view, the end-uses are essentially unchanged from the traditional goals of analytics, which includes correlation, root-cause analysis, predictive simulation, and triggers to name a few. Nevertheless, there are some defining characteristics that distinguish the Big Data challenge.

First of all, the use cases have evolved to include unstructured data, though still leveraging structured data in increasingly complex ways. In this way, Big Data can provide greater richness to enhance analytical capabilities. In addition, it can also provide more timely results. Secondly, a major change is the computer systems involved; dealing with a quantum of data several times larger than traditional systems have been dealing with necessitates the use of systems designed with parallelism as a core engineering principle, which is why traditional solutions may not always scale effectively. The key to success, however, may be somewhat removed from core engineering and data science algorithms. From Western Digital's experience, the key piece is data governance.

Data governance

Data governance is a rapidly evolving and poorly defined field. Nevertheless, having a data-driven culture from the outset, top-level management at Western Digital prioritised the creation of internal data governance structures. This was in response to increasing complaints from workers who were unable to find the data they needed to do their jobs. Most often this was not because the data was not captured and stored, but because there was no systematic approach to finding and understanding data assets scattered around the global organisation. Some specific issues were the lack of a standard naming convention, and lack of comprehensive metadata to describe the various attributes of individual pieces of data.

In response to this, a 'data council' was formed, comprising key representatives from business users and IT, which produced several beneficial outcomes, including a maintained data dictionary. Woo believes it was the involvement of end-users in the data council that led to its success. Drawing on the Western Digital experience, Woo argues that the key to success - regardless of the types of data sets involved - is to have data governance structures in place. To start, he recommends an organisation may implement a data governance structure for existing structured data sets (conventionally RDBMS systems). The next step would be to extend this data governance structure to include unstructured data (e.g. log files from equipment, as well as data captured from social networking services).

Resourcing and Malaysian environment

A wide skill set is required for an organisation to successfully harness Big Data, but Woo suggests that it must start with domain knowledge. It is crucial to be well grounded with the vision, goals, and objectives of the core business. Otherwise, all that data may simply prove too distracting to actually get anything relevant done. Woo observes, "One must know how to ask the

right questions." In addition to domain and business knowledge, a scientific background is helpful. "Big Data experts are few and far between". Given the steep learning curve, this scenario is unlikely to change very soon, and developing internal capacity is especially difficult. However, a core analytics team is one that is best grown at home, due to the need for a good grasp of evolving business imperatives. Infrastructure and technology aspects, on the other hand, are best outsourced to proven experts.

In general. Woo does not feel that Malaysian initiatives are particularly disadvantaged, because for the most part, the challenges around Big Data projects affect all companies worldwide. However, one specific frustration is the relative lack of relevant training. Woo believes that Malaysia is relatively underserved in this regard, compared to other locations such as Singapore.

"There is a clear need for independent benchmarking data in the Malaysian financial sector".

Ian Phoon, Malaysian bank

Ian Phoon works in risk management at a Malaysian bank assessing credit risk. Part of their department's role is assessing loans and setting limits for commercial and corporate loans. Phoon has worked at several local banks since graduating from the University of Malaya. He has a background in Management Information Systems including streams in computer science and user experience / usability. His four year degree program included one semester at the University of Melbourne in Australia, and an internship at a financial software business.

Data integrity and prediction

At his current employer, Phoon works with a criteria rating system. His department assesses the performance of clients in determining loan limits; their objective is to improve loan quality from a credit perspective. Phoon is responsible for data integrity, helping to analyse, prepare and visualise potential data problems. This involves preparing data, assessing data input quality and providing design feedback on backend systems. Big data offers opportunities for predictive analysis on prospects that are likely

to default through data modelling. The most valuable potential of Big Data is in predictive analysis for 'early warning systems'. This involves using historical information to determine patterns and triggers of company defaults on loans. External news reports and sentiment analysis from the market in assessing the performance of companies can provide valuable 'predictive' data.

Tools of the trade

In his current role, Phoon makes heavy use of traditional statistics tools with a long established track record in the finance domain, such as SAS and EViews. Nevertheless, he believes there is scope for new and improved tools and solutions (e.g. in areas such as data visualisation) and that the need for integration and interoperability between these tools is important. Despite the growing prevalence of analytics-as-a-service and other cloud-based Big Data solutions, cloud services is not likely to be adopted by the traditional banking industry anytime soon, particularly due to perceptions on security.

Phoon believes the most important skills for a data scientist are in managing databases and having expertise in statistics. General Business Intelligence skills are also important. He believes data scientists can help bridge the communication gap between business and IT functions. As a data scientist, data quality is a key concern in understanding how input integrity can affect prediction models. He believes there will be a growing demand for data scientist degrees, but first organisations need to understand the value that data scientists can add.

Industry benchmarking

Phoon says "there is a clear need for independent benchmarking data in the Malaysian financial sector". Current regulation prevents information sharing between financial institutions. Commercial banks would benefit from industry overviews such as the number of loans and customer segments as well as guidelines on best practice. Benchmarking data can also assist with predictive modelling. While newer business models such as online credit services are emerging, Phoon maintains that the customer-building strengths of traditional banks will give them a competitive edge.

In Malaysia, banking, telecommunications and marketing are important industries where Big Data can make an impact, but Phoon sees that adoption of Big Data technologies is not regarded as critical yet. He suggests that education around successes in specific sectors should be shared more widely, even with other industries, to help the rate of big data adoption. Campus talks from Big Data practitioners would also help spread awareness and drive the growth of talent supply.

Conclusion

In this report, we have sought to capture the state of Big Data analytics in Malaysia in all its diversity and varying levels of maturity. We know there is strong support for Big Data at grass-roots and executive levels. But aspirations and enthusiasm for Big Data is not uniformly matched with resource commitment. There are several inhibitors that prevent organisations in making headway with Big Data projects, including recruiting the appropriate skills, the ideation of novel, high value use cases and access to datasets. Our report raises many more questions. There are varying requirements for the elusive role of the data scientist. There are further questions of how Big Data budgets should be optimally split between infrastructure, data, human resources and other areas. We prompt stakeholders to consider:

- Are you committing resources aggressively enough?
- Are you driving a data-centric approach to all organisational processes?
- Are you considering out-of-the box end uses to give you a competitive edge?
- Are you prioritising the right blend of skills and capabilities?
- Are you participating in cultural change for Big Data advocacy?

Big Data involves not just technological evolution, but a cultural shift in the way we work and share knowledge. These values include an increasing emphasis on transparency and continuous innovation. The advent in Big Data brings rapid change but also emphasises interest in classical disciplines such as mathematics and statistics. The pitfalls in the use of Big Data should be at the forefront in considering adoption. Statisticians warn us that correlations do not equal causation, and that; the availability of increasing variables does not guarantee greater clarity. The privacy rights of individuals will also constrain how organisations collect, manage and distribute data. However, we believe the productive value of Big Data analytics in societies outweighs drawbacks if these concerns are respected and managed carefully. We have no doubt that the Big Data analytics industry will continue to mature and evolve in unexpected ways in Malaysia that will support the nation's competitiveness in the global economy.

About the authors

Sandra Hanchard

Sandra is the Organising chair of Big Data Malaysia. She is an industry analyst with expertise in consumer adoption of new technologies. Formerly, Sandra was the Asia Pacific lead for custom research with global Internet measurement firm, Experian Hitwise. She worked with leading blue-chip and new media brands to grow their businesses through insights derived from Big Data. Her commentary has featured prominently in the press in Australia. Currently, Sandra is the recipient of an Australian Postgraduate Award for her doctoral thesis which investigates social media information use in Malaysia. She is a past speaker at the Melbourne Writers Festival and a fellow of the Oxford Internet Institute Summer Doctoral Program at Oxford University.

sandra@bigdatamalaysia.org

Tirath Ramdas

Tirath founded Big Data Malaysia while working with Acunu, a London-based Big Data startup. Though that was his first role that explicitly wore the label 'Big Data', his first exposure to large scale information processing occurred many years earlier during his PhD in computer architecture for applications in computational quantum chemistry. Other prior roles include systems administration, embedded systems development for VoIP and video applications, non-profit strategy consulting, as well as business intelligence and data warehousing conceptual application design. Currently based primarily in Melbourne, Australia, he is a software development consultant to information security startup Bromium.

tirath@bigdatamalaysia.org

About Big Data Malaysia

The first Big Data London meetup was held in a co-working space near London's famous Leicester Square on the evening of May 25th 2011, featuring talks on Big Data use cases in interactive TV, games, and business intelligence. Founded by London-based startup Acunu, the group was created and developed to foster awareness and understanding of the potential for Big Data to transform businesses, industries, and society at large – while always digging into the technology itself.

One year later, the first Big Data Malaysia meetup was help in the Mercury room at Telekom Malaysia Research & Development in Cyberjaya on the afternoon of May 16th 2012, featuring talks on R, applications in genetics and air traffic modelling, as well as marketing services. Also founded by Acunu, Big Data Malaysia shared much of Big Data London's 'DNA' - in particular core values such as:

1. Maintaining a broad appeal (business executives in suits as welcome as hackers in shorts).
2. Embracing the wide spectrum of issues to do with Big Data (e.g. distributed systems and data journalism are equally relevant).
3. Emphasis on interesting content (more than just a networking forum).

Over time Big Data Malaysia has evolved in its own way to address needs and opportunities in Malaysia, embarking on exercises such as partnering with relevant organisations (such as community/ professional groups and commercial conference organisers), conducting a group brainstorming workshop to discuss the emerging impact of Big Data on society at large, coordinating Big Data Week 2013, and producing the "Big Data in Malaysia: Emerging Sector Profile" survey.

Although Big Data Malaysia has evolved to conduct activities beyond meetups, running regular meetups is a core aspect of the group, to allow people to network and learn. To that end in our first 24 months we have run many meetups as listed here:

- **"Hello World": May 2012**, sponsored by Acunu, hosted by Telekom Malaysia R&D, Cyberjaya, with talks by:
  - Julian Lee, Revolution Analytics
  - Daniel Walters, Experian Marketing Services
  - Hakim Albasrawy, Tandemic

- **"Bigger, Faster, Messier": July 2012**, sponsored by Acunu, hosted by iTrain, Kuala Lumpur, with talks by:
  - Nicolas Yip, Innity
  - Calum Halcrow and Kaveh Mousavi Zamani, RMG Technology

- **Joint meetup with SNIA Malaysia: September 2012**, sponsored by SNIA Malaysia and Acunu, hosted in Cyberjaya, with talks by:
  - SNIA tech tutorial on flash storage
  - Rachelle Foong, HP
  - Tirath Ramdas, Acunu

- **"Big Data in Telecommunications": October 2012**, organised and hosted by TM R&D, Cyberjaya with talks by:
  - Zach Tan, Google
  - Raymond Au, Oracle
  - Shamsul Anuar Yahaya, Telekom Malaysia

- **"Big Data in Social Media Analytics": December 2012**, hosted by iTrain, Kuala Lumpur, with talks by:
  - Sandra Hanchard, Swinburne Institute for Social Research

- **"Velvet Underground", 30th January 2013**, sponsored by Acunu and Experian, a networking event coinciding with the Market Research in the Mobile World (MRMW) conference, Zouk, Kuala Lumpur.

- **"Big Data in Market Research and Beyond", 1st February 2013**, sponsored by Acunu and Experian, coinciding with the Market Research in the Mobile World (MRMW) conference, hosted by iTrain, Kuala Lumpur, with talks by:
  - Colin Pal and Taufek Johar, Experian Marketing Services
  - Tirath Ramdas, Acunu
  - Shazri Shahrir, Telekom Malaysia R&D, and Nurazam Malim, Twistcode
  - Rebekah Lee, Integricity Interactive
  - Attendees also participated in a group brainstorming exercise, discussing how Big Data will impact society from the perspective of small business, big business, government, and individuals/consumers.

- **Malaysia was part of Big Data Week 2013 in mid-April 2013.** Big Data Malaysia assumed an oversight role during this festival, working to encourage a balanced program as well as fostering cross-promotion and minimising event overlaps. All together 9 separate events were part of the program, each of which was organised by different partner organisations:
  - TRANSIT: Distillation
  - Python Malaysia: Big Data Week Workshop
  - Big Data in Telecommunications 2
  - Big Data: Strategy, Practice, and Research
  - SNIA Malaysia Tech@Break
  - NoSQL Asia: Exploring The Technology Behind Big Data Week
  - GPU Technology Workshop Asia 2013
  - The Data Revolution: Powered by the AWS Cloud
  - Data Journalism: Storytelling in Malaysia with Big Data

- **"Big Data Malaysia - Anniversary 1.0": 30th May 2013**, hosted by iTrain, Kuala Lumpur, Emceed by Daniel Walters, Experian, with talks by:
  - Khoo Swee Chuan, Netapp
  - Siah Hark Cheong and Mun Heng, from a multinational financial services company, but speaking in their personal capacity

- **"Big Data Malaysia, with F-Secure": 1st August 2013**, sponsored and hosted by F-Secure, Kuala Lumpur, emceed by Ng Swee Meng, OnApp, with talks by:
  - Keng Chee Chan, F-Secure
  - Aizu Ikmal Ahmad, Malaysiakini
  - Zakiullah Khan Mohammed, independent software engineering management consultant

- **"Big Data in Telecommunications 3": 24th September 2013**, sponsored and hosted by TM R&D, Cyberjaya, emceed by Tirath Ramdas, Marklar Marklar Consulting, with talks by:
  - Mohamad Ansahari Bin Abdul Kudus, Telekom Malaysia
  - Szilard Barany, Teradata
  - Rosmawati Binti Ab Raub, Telekom Malaysia Research & Development

- **"Cents & Security": 20th February 2014**, sponsored and hosted by F-Secure, Bangsar South, emceed by Szilard Barany, Teradata, with talks by:
  - Foong Chee Mun, MoneyLion
  - Goh Su Gim, F-Secure
  - Pak Mei Yuet, MDeC

The first two years of Big Data Malaysia have uncovered a significant and still rapidly growing interest and excitement for the possibilities of Big Data in Malaysia, and we look forward to seeing what comes next. We believe that a year from now, we shall look back to see that this was the turning point where the majority of discussion changes from "this is what we could do" to "this is what we're doing."

Big Data Malaysia exists on a number of social media sites; our Facebook group is especially active, and we also have a LinkedIn group, a Meetup.com group, and a mailing list. Links to all these properties may be found on our landing page at http://bigdatamalaysia.org